Math 170S: Homework 1

Owen Jones

10/13/2023

Problem 1.

1. $\overline{x} = 7.2333, s^2 = 4.1823, s = 2.0451$
2. within one std: 24, within two std: 29

Problem 2. $\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{\sum_{i=1}^{n} ax_i + b}{n} = \frac{nb + a\sum_{i=1}^{n} x_i}{n} = a\frac{\sum_{i=1}^{n} x_1}{n} + b = a\overline{x} + b$

$$s_y^2 = \frac{\sum_{i=1}^{n} (y_i - \overline{y})^2}{n-1} = \frac{\sum_{i=1}^{n} (ax_i + b - (a\overline{x} + b))^2}{n-1} = \frac{a^2\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1} = a^2 s_x^2$$

Problem 3.

1. $\pi_{25} = 6.0, \pi_{75} = 8.25, IQR = 2.25$
2. $\pi_{10} = 5.18, \pi_{90} = 9.34$
3. suspected outliers: 14.1

Problem 4.

1. $P(Y_7 < 27.3) = \sum_{k=7}^{8} \binom{8}{k} (0.7)^k (0.3)^{8-k} = 0.2553$
2. $P(Y_5 < 27.3 < Y_8) = \sum_{k=5}^{7} \binom{8}{k} (0.7)^k (0.3)^{8-k} = 0.7482$

```
In [1]:  import numpy as np
         import pandas as pd
         import math

In [2]:  lead_concentrations = np.array([6.7, 5.4, 5.2, 6.0, 8.7,
                                       6.0, 6.4, 8.3, 5.3, 5.9,
                                       7.6, 5.0, 6.9, 6.8, 4.9,
                                       6.3, 5.0, 6.0, 7.2, 8.0,
                                       8.1, 7.2, 10.9, 9.2, 8.6,
                                       6.2, 6.1, 14.1, 10.6, 8.4])
         lead_concentrations

Out[2]:  array([ 6.7,  5.4,  5.2,  6. ,  8.7,  6. ,  6.4,  8.3,  5.3,  5.9,  7.6,
                 5. ,  6.9,  6.8,  4.9,  6.3,  5. ,  6. ,  7.2,  8. ,  8.1,  7.2,
                10.9,  9.2,  8.6,  6.2,  6.1, 14.1, 10.6,  8.4])

In [3]:  lead_mean=np.mean(lead_concentrations)
         lead_mean

Out[3]:  7.2333333333333325

In [4]:  lead_std=np.std(lead_concentrations,ddof=1)
         lead_std

Out[4]:  2.0450669550346543

In [5]:  lead_var=np.var(lead_concentrations,ddof=1)
         lead_var

Out[5]:  4.182298850574712

In [6]:  a=lead_mean+lead_std

In [7]:  b=lead_mean-lead_std

In [8]:  within_one_std=lead_concentrations[(lead_concentrations<=a) & (lead_concentrations>=b)]
         within_one_std.size

Out[8]:  24

In [9]:  a_2=lead_mean+2*lead_std
         b_2=lead_mean-2*lead_std

In [10]: within_two_std=lead_concentrations[(lead_concentrations<=a_2) & (lead_concentrations>=b_2)]
         within_two_std.size

Out[10]: 29

In [11]: pi_25=np.percentile(lead_concentrations,25)
         pi_25

Out[11]: 6.0

In [12]: pi_75=np.percentile(lead_concentrations,75)
         pi_75

Out[12]: 8.25

In [13]: IQR=pi_75-pi_25
         IQR

Out[13]: 2.25

In [14]: pi_10=np.percentile(lead_concentrations,10)
         pi_10

Out[14]: 5.180000000000001

In [15]: pi_90=np.percentile(lead_concentrations,90)
         pi_90

Out[15]: 9.340000000000002

In [16]: suspected_outliers=lead_concentrations[(lead_concentrations<=pi_25-1.5*IQR) | (lead_concentrations>=pi_75+1.5*IQR)]
         suspected_outliers

Out[16]: array([14.1])

In [17]: prob_seventh_order_stat=math.comb(8,7)*(0.7)**7*(0.3)+math.comb(8,8)*(0.7)**8
         prob_seventh_order_stat

Out[17]: 0.2552983299999999

In [18]: prob_seventh_order_stat_2=math.comb(8,7)*(0.7)**7*(0.3)+math.comb(8,6)*(0.7)**6*(0.3)**2+math.comb(8,5)*(0.7)**5*(0.3)**3
         prob_seventh_order_stat_2

Out[18]: 0.7482476399999998

In [ ]:
```

Problem 5.

1. $E[W_r^2] = \int_0^1 w^2 g_r(w)dw$
$$= \int_0^1 w^2 \frac{n!}{(r-1)!(n-r)!}[w]^{r-1}[1-w]^{n-r}dw$$
$$= \frac{r(r+1)}{(n+1)(n+2)} \int_0^1 \frac{(n+2)!}{(r+1)!(n-r)!}[w]^{r+1}[1-w]^{n-r}dw$$
$$= \frac{r(r+1)}{(n+1)(n+2)} \int_0^1 Beta(r+2, n-r+1)dw$$
$$= \frac{r(r+1)}{(n+1)(n+2)} \cdot 1$$

2. $Var[W_r] = E[W_r^2] - E[W_r]^2 = \frac{r(r+1)}{(n+1)(n+2)} - \frac{r^2}{(n+1)^2} = \frac{r(r+1)(n+1)}{(n+1)^2(n+2)} - \frac{r^2(n+2)}{(n+1)^2(n+2)} = \frac{r(n-r+1)}{(n+1)^2(n+2)}$